

GenCore version 5.1.7  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.  
 OM protein - protein search, using sw model  
 Run on: February 14, 2006, 17:53:18 ; Search time 45 Seconds  
 (without alignments)  
 1821.704 Million cell updates/sec  
 Title: US-10-612-012-2  
 Perfect score: 852  
 Sequence: 1 MSSVSPQIPSRPLLLTHE.....AAFDGGFTVTRPGLNSKL 852  
 Scoring table: OLIGO  
 Gapop 60.0 , Gapext 60.0  
 Searched: 283416 seqs, 96216763 residues  
 Word size : 10

Total number of hits satisfying chosen parameters: 50

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 80.\*

1: Pir1.\*  
 2: Pir2.\*  
 3: Pir3.\*  
 4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	2.0	774	2 B84031	ATP-dependent prot
2	17	2.0	795	1 A70322	endopeptidase La (
3	16	1.9	784	2 AC0383	endopeptidase La (
4	16	1.9	875	2 T09142	endopeptidase La (
5	15	1.8	779	1 B42375	endopeptidase La h
6	14	1.6	798	2 G83420	Lon proteinase PAL
7	14	1.6	827	1 A36895	endopeptidase La (
8	14	1.6	848	2 C82712	ATP-dependent seri
9	13	1.5	773	2 T27382	hypothetical prote
10	13	1.5	778	2 E97224	ATP-dependent Lon
11	13	1.5	786	2 A96956	ATP-dependent prot
12	13	1.5	799	2 F83549	probable ATP-depen
13	13	1.5	817	1 A49844	endopeptidase La (
14	13	1.5	821	2 B75530	ATP-dependent prot
15	13	1.5	885	1 T04321	endopeptidase La h
16	13	1.5	954	1 T04325	probable ATP-depen
17	12	1.4	784	1 S47270	endopeptidase La (
18	12	1.4	784	1 SUECLA	endopeptidase La*
19	12	1.4	784	2 AE0558	Lon protease (impo
20	12	1.4	784	2 E90690	endopeptidase La (
21	12	1.4	787	2 A72230	endopeptidase La (
22	12	1.4	799	2 A85541	hypothetical prote
23	12	1.4	805	2 AH2731	ATP-dependent prot
24	12	1.4	805	2 G97512	ATP-dependent prot
25	12	1.4	810	1 JC6045	endopeptidase La (
26	12	1.4	820	2 H81106	ATP-dependent prot
27	12	1.4	820	2 H81908	probable endopepti
28	12	1.4	881	2 B71316	endopeptidase La (
29	12	1.4	1067	1 S62421	endopeptidase La h

## ALIGNMENTS

## RESULT 1

B84031

ATP-dependent proteinase La 1 (lon) (class III heat-shock protein) lonA [imported] - Ba  
 C:Species: Bacillus halodurans  
 C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
 C:Accession: B84031

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, P.; Hir  
 Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84031

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-774 &lt;STO&gt;

A:Cross-references: UNIPROT:Q9K8F6; UNIPARC:UPI00000C40AD; GB:AP001517; GB:BA000004; NFI

A:Experimental source: strain C-125

C:Genetics:

A:Gene: lonA

C:Superfamily: ATP-dependent Lon protease

Query Match Best Local Similarity 2.0%; Score 17; DB 2; Length 774;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 453 GDPAAALLEVLDPEQNH 469  
 |||||  
 Db 433 GDPAAALLEVLDPEQNH 449

## RESULT 2

A70322

endopeptidase La (EC 3.4.21.53) - Aquifex aeolicus  
 N:Alternate names: ATP-dependent proteinase lon; ATP-dependent serine proteinase La  
 N:Contains: adenosinetriphosphatase (EC 3.6.1.3)  
 C:Species: Aquifex aeolicus  
 C>Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 09-Jul-2004  
 C:Accession: A70322

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
 Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: A70322  
 A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-795 <AOF>

A:Cross-references: UNIPROT:O66605; UNIPARC:UPI00000562D8; GB:AE000680; NID:g2982948; PI  
 A:Experimental source: strain VP5

C:Comment: This allosteric enzyme catalyzes the hydrolysis of large proteins in the pres  
 C:Genetics:  
 A:Gene: lon

C:Superfamily: ATP-dependent Lon protease

12 1.4 1096 1 T01765 endopeptidase La-1  
 31 12 1133 1 S43938 endopeptidase La-1  
 32 11 779 2 E97778 endopeptidase La (

33 11 784 1 A71704 endopeptidase La (

34 11 786 2 G82141 ATP-dependent LA p

35 11 1.3 795 1 D64226 endopeptidase La (

36 11 1.3 803 1 A64070 endopeptidase La (

37 11 1.3 813 1 D70176 endopeptidase La (

38 11 1.3 819 1 B72128 endopeptidase La (

39 11 1.3 819 1 C71527 Lon ATP-dependent

40 11 1.3 819 2 E86494 proteinase, Lon fa

41 11 1.3 819 2 E81681 endopeptidase La (

42 10 1.2 774 1 I40421 endopeptidase La (

43 10 1.2 777 2 F84985 ATP-dependent prot

44 10 1.2 791 2 D82901 endopeptidase La (

45 10 1.2 795 1 S73830 endopeptidase La (

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2006, 17:49:34 ; Search time 189 Seconds  
(without alignments)  
1980.692 Million cell updates/sec

Title: US-10-612-012-2  
Perfect score: 852  
Sequence: 1 MSSVSPQIPSRPLLLTHE.....AAPDGGFTVKTPLGSLSKL 852

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 10

Total number of hits satisfying chosen parameters: 149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	852	5	ABG32363 Human ATP
2	852	100.0	852	7	Adm56266 Human ATP
3	852	100.0	852	9	Adv85873 Human pro
4	815	95.7	852	4	Aab47562 Protease
5	815	95.7	852	7	Adc10040 Human NOV
6	714	83.8	852	4	Aam93785 Human pol
7	714	83.8	852	5	Abp69443 Human pro
8	714	83.8	852	8	Adl31771 Human pro
9	693	81.3	693	7	Adj70423 Human hea
10	581	68.2	581	4	Aab95332 Human pro
11	581	68.2	581	4	AAB99179 Human pro
12	470	55.2	571	4	AAB94288 Human pro
13	432	50.7	432	4	AAE11085 Human pro
14	432	50.7	432	4	AAE11085 Human pro
15	421	49.4	433	6	ABU11611 Human MDD
16	379	44.5	581	8	Adr08963 Human pro
17	331	38.8	423	7	AAU17375 Human pro
18	331	38.8	423	7	ADP94083 Human nov
19	215	25.2	228	4	AAU17371 Human sig
20	215	25.2	228	7	ADP94079 Human nov
21	195	22.9	391	5	ABB89248 Human pol
22	160	18.8	160	8	ADP55489 Human PRO
23	118	13.8	118	4	AAE01658 Human gen
24	85	10.0	85	8	ABO57485 Human gen

25	78	9.2	78	4	AAm14167	AAm14167 Peptide #
26	78	9.2	78	4	ABB33114	ABB33114 Peptide #
27	78	9.2	78	4	AAm26575	AAm26575 Peptide #
28	78	9.2	78	4	ABB27940	ABB27940 Human pep
29	78	9.2	78	4	ABB18580	ABB18580 Protein #
30	78	9.2	78	4	ABG47963	ABG47963 Human liv
31	78	9.2	78	4	AAm01905	AAm01905 Peptide #
32	78	9.2	78	5	ABG35946	ABG35946 Human pep
33	50	5.9	50	8	ABO57622	ABO57622 Human gen
34	32	3.8	102	5	ABP07459	ABP07459 Human ORF
35	31	3.6	101	4	AAm91575	AAm91575 Human imm
36	21	2.5	62	4	AAU32043	AAU32043 Novel hum
37	17	2.0	774	8	ADS28353	ADS28353 Bacterial
38	17	2.0	786	8	ADN26262	ADN26262 Bacterial
39	17	2.0	795	8	ADN17367	ADN17367 Bacterial
40	16	1.9	767	8	ADN27141	ADN27141 Bacterial
41	16	1.9	777	8	ADN26487	ADN26487 Bacterial
42	16	1.9	784	6	ABU50220	ABU50220 Protein e
43	16	1.9	785	6	ABM66958	ABM66958 Photorhab
44	16	1.9	795	8	ADS42793	ADS42793 Bacterial
45	16	1.9	892	8	ADN20874	ADN20874 Bacterial

## ALIGNMENTS

RESULT 1  
ABG32363  
ID ABG32363 standard; protein; 852 AA.  
XX AC ABG32363;  
XX DT 15-NOV-2002 (first entry)  
XX DE Human ATP-dependent protease.  
XX KW Human; enzyme; ATP-dependent protease; developmental disorder; cancer;  
KW retinoblastoma; melanotic melanoma; endometrial adenocarcinoma;  
KW ovarian adenocarcinoma; schizophrania.  
XX OS Homo sapiens.  
XX PN US2002081704-A1.  
XX PD 27-JUN-2002.  
XX PF 21-DEC-2000; 2000US-00741150.  
XX PR 22-NOV-2000; 2000US-0252410P.  
XX PA (GUEG/) GUEGLER K.  
XX PA (WEBS/) WEBSTER M.  
XX PA (YANC/) YAN C.  
XX PA (SHAO/) SHAO W.  
XX PA (KETC/) KETCHUM K A.  
XX PA (DFRA/) DI FRANCESCO V.  
XX PA (BEAS/) BEASLEY E M.  
XX FI Guegler K, Webster M, Yan C, Shao W, Ketchum KA, Di Francesco V;  
XX PI Beasley EM;  
XX WPI: 2002-635469/68.  
XX N-PSDB: ABK90887, ABK90888.  
XX DR New human ATP-dependent peptides and encoding nucleic acids, useful for  
XX PT diagnosing, preventing and/or treating disorders like cancer, e.g.  
XX PT retinoblastomas, melanomas, endometrial and ovarian adenocarcinomas, and  
XX PS schizophrania.  
XX CC Claim 1; Fig 2; 119pp; English.  
XX The invention relates to a new isolated ATP-dependent protease (I) and  
CC the nucleic acid encoding it. The peptides are useful in identifying

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2006, 17:58:02 ; Search time 169 Seconds  
(without alignments)

2106.453 Million cell updates/sec

Title: US-10-612-012-2

Perfect score: 852

Sequence: 1 MSSVSPFIQPSRLPLLLTHE.....AAFDGGFTVKTTPGLNSKL 852

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 10

Total number of hits satisfying chosen parameters: 124

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pgp:\*
- 2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pgp:\*
- 3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pgp:\*
- 4: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB.pgp:\*
- 5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pgp:\*
- 6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pgp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	852	100.0	852	3	US-09-741-150-2
2	852	100.0	852	5	US-10-612-012-2
3	815	95.7	852	5	US-10-239-032-4
4	693	81.3	693	4	US-10-408-765A-2229
5	432	50.7	432	4	US-10-156-239-44
6	331	38.8	423	3	US-09-764-868-940
7	215	25.2	228	3	US-09-764-868-936
8	195	22.9	391	4	US-10-264-237-1624
9	85	10.0	85	4	US-10-029-386-31119
10	78	9.2	78	3	US-09-864-761-33878
11	50	5.9	50	4	US-10-029-386-31256
12	17	2.0	774	4	US-10-369-493-17386
13	17	2.0	786	4	US-10-369-493-8915
14	17	2.0	795	4	US-10-369-493-20
15	16	1.9	670	4	US-10-424-599-183947
16	16	1.9	767	4	US-10-369-493-9794
17	16	1.9	777	4	US-10-369-493-9140
18	16	1.9	784	4	US-10-282-122A-78144
19	16	1.9	795	4	US-10-369-493-21223
20	16	1.9	892	4	US-10-369-493-3527
21	15	1.8	280	4	US-10-425-114-43912
22	15	1.8	376	4	US-10-424-599-198007
23	15	1.8	766	4	US-10-369-493-16141
24	15	1.8	768	4	US-10-282-122A-45835
25	15	1.8	769	4	US-10-369-493-16622
26	15	1.8	777	4	US-10-369-493-373
27	15	1.8	784	4	US-10-282-122A-68623

#### ALIGNMENTS

##### RESULT 1

US-09-741-150-2  
; Sequence 2, Application US/09741150  
; Publication No. US20020081704A1  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: NUCLES THEREOF  
; FILE REFERENCE: CL000968  
; CURRENT APPLICATION NUMBER: US/09/741,150  
; CURRENT FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 852  
; TYPE: PRT  
; ORGANISM: Human  
US-09-741-150-2

Query Match 100.0%; Score 852; DB 3; Length 852;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSSVSPFIQPSRLPLLLTHEGVLLPGSTWRTSVDSAHNLQVRSRLKGTSLQSTILGVI	60
DB	1	MSSVSPFIQPSRLPLLLTHEGVLLPGSTWRTSVDSAHNLQVRSRLKGTSLQSTILGVI	60
QY	61	PNTPDPSAQDLPLHRIQTALAVQVVGSNWPKPHYTLITGLCRFQIVQVLEKPY	120
DB	61	PNTPDPSAQDLPLHRIQTALAVQVVGSNWPKPHYTLITGLCRFQIVQVLEKPY	120
QY	121	IAEVEQLDRLEFPNTCKRRELGESEQFYKAVQVLEMDMSVPAVAKRLDLSLR	180
DB	121	IAEVEQLDRLEFPNTCKRRELGESEQFYKAVQVLEMDMSVPAVAKRLDLSLR	180
QY	181	EALPDILTITTSNKEKIQLDAVLEERFKMTIPLVROHGLKLLQKTRKPKQDDK	240
DB	181	EALPDILTITTSNKEKIQLDAVLEERFKMTIPLVROHGLKLLQKTRKPKQDDK	240
QY	241	RVIATPIRTHISCTLEDEDEDDNDIVLEKKIRTSSMPEQAHKVCVKIKRLKGM	300
DB	241	RVIATPIRTHISCTLEDEDEDDNDIVLEKKIRTSSMPEQAHKVCVKIKRLKGM	300
QY	301	QOSMPEYALTRYLEMLVLPWNKSTTDRLDRAARILLDNDHYAMEKLRVLEYLAVR	360
DB	301	QOSMPEYALTRYLEMLVLPWNKSTTDRLDRAARILLDNDHYAMEKLRVLEYLAVR	360
QY	361	QLKNNLKGPILCFVGPVGVKTSVGRSVAKTLGRBFHRIALGVQDSODIRGRHRTVGS	420
DB	361	QLKNNLKGPILCFVGPVGVKTSVGRSVAKTLGRBFHRIALGVQDSODIRGRHRTVGS	420

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2006, 17:58:56 ; Search time 17 Seconds  
(without alignments)  
657.671 Million cell updates/sec

Title: US-10-612-012-2

Perfect score: 852

Sequence: 1 MSSVSPQIPSRPLLLTHE.....AAFDGGFTVKTTPGLNSKL 852

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 97014 seqs, 13122538 residues

Word size : 10

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA New:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08 NEW PUB.pap.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06 NEW PUB.pap.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07 NEW PUB.pap.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW PUB.pap.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09 NEW PUB.pap.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10 NEW PUB.pap.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11 NEW PUB.pap.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60 NEW PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.5	830	7 US-11-098-686-10393	Sequence 10393, A
2	12	1.4	804	7 US-11-070-080-18	Sequence 18, Appl
3	12	1.4	820	6 US-10-467-657-4910	Sequence 4910, Ap
4	11	1.3	817	7 US-11-098-686-11041	Sequence 11041, A

#### ALIGNMENTS

RESULT 1  
US-11-098-686-10393  
; Sequence 10393, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10393  
; LENGTH: 830  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10393

Query Match 1.5%; Score 13; DB 7; Length 830;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 LTGQLGDVWKESA 697

Db 667 LTGQLGDVWKESA 679

#### RESULT 2

US-11-070-080-18  
; Sequence 18, Application US/11070080  
; Publication No. US20050287625A1  
; GENERAL INFORMATION:  
; APPLICANT: Ye, Rick W.

; TITLE OF INVENTION: PROCESS FOR EXPRESSION OF FOREIGN GENES IN METHANE METABOLIZING  
; TITLE OF INVENTION: BACTERIA THROUGH CHROMOSOMAL INTEGRATION  
; FILE REFERENCE: CL-2443 US NA  
; CURRENT APPLICATION NUMBER: US/11/070,080  
; CURRENT FILING DATE: 2005-03-02  
; PRIOR APPLICATION NUMBER: US 60/550385  
; PRIOR FILING DATE: 2004-03-05  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 18  
; LENGTH: 804  
; TYPE: PRT  
; ORGANISM: Methylobomonas sp. 16a  
US-11-070-080-18

Query Match 1.4%; Score 12; DB 7; Length 804;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 458 ALLEVLDPQNH 469

Db 440 ALLEVLDPQNH 451

#### RESULT 3

US-10-467-657-4910  
; Sequence 4910, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA

; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 4910  
; LENGTH: 820  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-4910

Query Match 1.4%; Score 12; DB 6; Length 820;  
Best Local Similarity 100.0%; Pred. No. 0.0012;

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2006, 17:49:51 / Search time 238 Seconds  
(without alignments)  
2525.673 Million cell updates/sec

Title: US-10-612-012-2  
Perfect score: 852  
Sequence: 1 MSSVSPQIPRLPLLLTHE.....AAPDGGFTVKTPLGLLSKL 852

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 10

Total number of hits satisfying chosen parameters: 265

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	815	95.7	852	2	Q86W48	homo sapien
2	715	83.9	715	2	Q8N3B9	homo sapien
3	714	83.8	852	2	Q8NCE9	homo sapien
4	581	68.2	581	2	Q96K43	homo sapien
5	273	32.0	852	2	Q8R6M5	PONPY
6	204	23.9	204	2	Q9BU35	HUMAN
7	202	23.7	228	2	Q8ZMF9	HUMAN
8	147	17.3	806	2	Q6TX13	RAT
9	147	17.3	852	2	Q9DBN5	MOUSE
10	85	10.0	160	2	Q9D1A6	MOUSE
11	85	10.0	432	2	Q8BK00	MOUSE
12	79	9.3	840	2	Q5PQV6	BRARE
13	69	8.1	114	2	Q8NF29	HUMAN
14	30	3.5	874	2	Q4T681	TETNG
15	21	2.5	222	2	Q9XKK8	LITER
16	18	2.1	489	2	Q6AJ30	DESPS
17	18	2.1	774	2	Q8AK61	DESPS
18	17	2.0	768	2	Q74EN9	GEOSL
19	17	2.0	774	2	Q9K8F6	BACHD
20	17	2.0	795	1	LON AQUAE	
21	17	2.0	795	2	Q9LCK1	THETH
22	17	2.0	795	2	Q72KS4	THET2
23	17	2.0	795	2	Q5SK78	THET6
24	17	2.0	817	2	Q83DU0	COXBU
25	17	2.0	819	2	Q747S2	GEOSL
26	17	2.0	826	2	Q7NBA5	MYCGA
27	16	1.9	784	2	Q66DT2	VERPS
28	16	1.9	784	2	Q8ZC57	VERPE
29	16	1.9	784	2	Q7N0L5	PHOLL
30	16	1.9	793	2	Q6D825	ERWCT
31	16	1.9	802	2	Q8D154	YERSINIA PE

32 16 1.9 812 2 Q4NQW6\_9DELT  
33 16 1.9 843 2 Q4NQW6\_9DELT  
34 16 1.9 875 1 LONH1\_SPTOL  
35 16 1.9 937 2 Q7SA85\_NEUCR  
36 16 1.9 1147 2 Q6BUJ8\_DEBHA  
37 15 1.8 557 2 Q8GT60\_ARATH  
38 15 1.8 702 2 Q725X1\_DESVH  
39 15 1.8 773 2 Q4MS75\_BACCE  
40 15 1.8 773 2 Q72ZV6\_BACCI  
41 15 1.8 773 2 Q81LC1\_BACAN  
42 15 1.8 774 2 Q5OXP0\_IDILO  
43 15 1.8 776 2 Q6HSS4\_BACAN  
44 15 1.8 776 2 Q817Q4\_BACR  
45 15 1.8 776 2 Q6HD56\_BACHK

## ALIGNMENTS

## RESULT 1

Q86W48\_HUMAN  
ID Q86W48\_HUMAN PRELIMINARY; PRT; 852 AA.  
AC Q86W48;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)  
DE Peroxisomal lon protease.  
GN Name=LONP;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA De Walque S., Van Veldhoven P.P.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uedini T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Blakesley R.W., Young A.C., Shevchenko Y., Souffard G.G.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schemm A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Brain;  
RX NIH MGC Project;  
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
EMBL; AJ548761; CAD68987.1; -; mRNA.  
EMBL; BC093912; AAH93912.1; -; mRNA.  
EMBL; BC093910; AAH93910.1; -; mRNA.  
DR HSSP; P08177; IRR9.  
DR MEROPS; S16.006; -.  
DR Ensembl; ENSG00000102910; Homo sapiens.  
DR GO; GO:0005524; P:ATP binding; IEA.

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 14, 2006, 17:57:16 ; Search time 49 Seconds

(without alignments)  
1437.544 Million cell updates/sec

Title: US-10-612-012-2

Perfect score: 852

Sequence: 1 MSSVSPIQIPSRPLLLTHE.....AAPDGGFTVKTRPGLLSKL 852.

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 10

Total number of hits satisfying chosen parameters: 16

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	852	2	US-09-741-150-2
2	852	100.0	852	2	US-10-160-187-2
3	15	1.8	790	2	US-09-543-681A-6059
4	14	1.6	810	2	US-09-902-540-11337
5	14	1.6	820	2	US-09-252-931A-30204
6	13	1.5	809	2	US-09-252-931A-29280
7	13	1.5	884	2	US-09-741-150-4
8	13	1.5	884	2	US-10-160-187-4
9	12	1.4	783	2	US-09-248-796A-16773
10	12	1.4	845	2	US-09-489-039A-13590
11	11	1.3	440	2	US-09-198-452A-44
12	11	1.3	825	2	US-09-438-185A-29
13	10	1.2	157	2	US-09-861-451A-8
14	10	1.2	424	2	US-09-198-452A-45
15	10	1.2	822	2	US-09-328-352-5754
16	10	1.2	832	2	US-09-540-236-3056

ALIGNMENTS

RESULT 1  
US-09-741-150-2  
; Sequence 2, Application US/09741150  
; Patent No. 6436689  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al

; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; FILE REFERENCE: CL000968  
; CURRENT APPLICATION NUMBER: US/09/741,150  
; CURRENT FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 852  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-741-150-2

Query Match 100.0%; Score 852; DB 2; Length 852;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSSVSPIQIPSRPLLLTHEGVLLPGSTWRTSVDSAHNLQVRSRLKGTSLQSTILGVI	60
Db	1	MSSVSPIQIPSRPLLLTHEGVLLPGSTWRTSVDSAHNLQVRSRLKGTSLQSTILGVI	60
Qy	61	PNTDPASDAQDLPPLHRIQTAALAVQVVGSNWPKPHYTLITGLCRFQIVQVLEKPY	120
Db	61	PNTDPASDAQDLPPLHRIQTAALAVQVVGSNWPKPHYTLITGLCRFQIVQVLEKPY	120
Qy	121	IAEVEQLDRLEFPNTCKMREELGELSEQFYKAVQVLEMDMSVPAVKLRLLDSLR	180
Db	121	IAEVEQLDRLEFPNTCKMREELGELSEQFYKAVQVLEMDMSVPAVKLRLLDSLR	180
Qy	181	EALPDILTSIIRTSNKEKQILDAVLEERFKMTIPLLVRQIEGLKLLQTRKPKQDDK	240
Db	181	EALPDILTSIIRTSNKEKQILDAVLEERFKMTIPLLVRQIEGLKLLQTRKPKQDDK	240
Qy	241	RVIAIRPIRITHISCTLEDEDEDEDNDIVLEKKIRTSMSPEQAHKVCVKEIKRLKM	300
Db	241	RVIAIRPIRITHISCTLEDEDEDEDNDIVLEKKIRTSMSPEQAHKVCVKEIKRLKM	300
Qy	301	PQSMPEYALTRYLELWVLPWNKSTTDRLDIRAARILLDNDHYAMEKLRVLEYLAVR	360
Db	301	PQSMPEYALTRYLELWVLPWNKSTTDRLDIRAARILLDNDHYAMEKLRVLEYLAVR	360
Qy	361	QLKNNLKGPILCFVGPVGKTSVGRSVAKTLGRFHRIRALGGVCDOSDIRHRTTVGS	420
Db	361	QLKNNLKGPILCFVGPVGKTSVGRSVAKTLGRFHRIRALGGVCDOSDIRHRTTVGS	420
Qy	421	MPGRIINGLKTGVNPNVFLLDVGLGKSLQGDPAALAEVLDPENHNFTHLYLVAF	480
Db	421	MPGRIINGLKTGVNPNVFLLDVGLGKSLQGDPAALAEVLDPENHNFTHLYLVAF	480
Qy	481	DLSQLFIATANTTATIPALLDRMEIIQVPGYTQBEKIBIAHRHLIPKLEHQHGLTPQ	540
Db	481	DLSQLFIATANTTATIPALLDRMEIIQVPGYTQBEKIBIAHRHLIPKLEHQHGLTPQ	540
Qy	541	IQIPQVTTLDIITRYTREAGVRSILDKLGAICRAVAVKVAEGQHKAKLDRSDVTERGC	600
Db	541	IQIPQVTTLDIITRYTREAGVRSILDKLGAICRAVAVKVAEGQHKAKLDRSDVTERGC	600
Qy	601	REHILEDEKPEISDITTDALPPMPFLIDPHALKDILGPPMYEMEVSQRLSQPGVALGL	660
Db	601	REHILEDEKPEISDITTDALPPMPFLIDPHALKDILGPPMYEMEVSQRLSQPGVALGL	660
Qy	661	AWTPLGGEIMFVBSRMDGEGQLTLTGQLGDMVKESAHLAISMLRSNAKYYQLTNAFGSF	720
Db	661	AWTPLGGEIMFVBSRMDGEGQLTLTGQLGDMVKESAHLAISMLRSNAKYYQLTNAFGSF	720
Qy	721	DLLENTDILHLFPAGAVTKDGPAGVTIVTCLASLFSGLRVSRSDVAMTGEITLGLVLPV	780
Db	721	DLLENTDILHLFPAGAVTKDGPAGVTIVTCLASLFSGLRVSRSDVAMTGEITLGLVLPV	780
Qy	781	GGIKDKVLAHRAAGLKVIIIPRNEKDLGIPGNVRQDLSFVTASCLDEVLNAAFDGSGT	840
Db	781	GGIKDKVLAHRAAGLKVIIIPRNEKDLGIPGNVRQDLSFVTASCLDEVLNAAFDGSGT	840